

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT:
 - (A) NAME: von Schaewen, Antje Dr. rer. nat.
 - (B) STREET: Natruperstrasse 169a
 - (C) CITY: Osnabrueck
 - (E) COUNTRY: Germany
 - (F) POSTAL CODE (ZIP): D-49076
 - (G) TELEPHONE: +49-541-684029
- (ii) TITLE OF INVENTION: Plant gntI sequences and the use thereof for the production of plants having reduced or lacking N-acetyl glucosaminyl transferase I (GnTI) activity
- (iii) NUMBER OF SEQUENCES: 6
- (iv) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1669 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA to mRNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Solanum tuberosum
 - (B) STRAIN: Desiree
 - (D) DEVELOPMENTAL STAGE: Sink organ
 - (F) TISSUE TYPE: Mesophyll
 - (G) CELL TYPE: Leaf cells
- (vii) IMMEDIATE SOURCE:
 - (A) LIBRARY: Lambda ZAP II (Eco RI)
 - (B) CLONE: gntI-A1(K)

(ix) FEATURE:

- (A) NAME/KEY: misc feature
- (B) LOCATION: 659..667
- (D) OTHER INFORMATION: /function= "Asn codon in this context is a potential glycosylation site"
- /product= "N-glycosylation consensus sequence"
- /phenotype= "N-glycans modulate protein properties"
- /standard_name= "N-glycosylation site"
- /label= pot-CHO
- /note= "GnTI-coding sequences from animals do not contain this feature"

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION:53..1393
- (C) IDENTIFICATION METHOD: experimental
- (D) OTHER INFORMATION:/codon_start= 53
 - /function= "initiates complex N-glycans on secretory glycoproteins"
 - /EC_number= 2.4.1.101
 - /product= "beta-1,2-N-acetylglucosaminyltransferase I"
 - /evidence= EXPERIMENTAL
 - /gene= "cgl"
 - /standard_name= "gntI"
 - /label= ORF
 - /note= "first gntI sequence from potato (unpublished)"

(ix) FEATURE:

- (A) NAME/KEY: 5'UTR
- (B) LOCATION:15..52

(ix) FEATURE:

- (A) NAME/KEY: 3'UTR
- (B) LOCATION:1394..1655

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION:80..139
- (D) OTHER INFORMATION:/function= "membrane anchor (amino acids 10-29)"
 - /product= "hydrophobic amino acid stretch in GntI"
 - /standard_name= "membrane anchor of a type II Golgi protein"
 - /note= "identified by comparison with GntI sequences from animals"

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION:1..14
- (D) OTHER INFORMATION:/function= "used for cloning the cDNA library in Lambda ZAPII"
 - /product= "EcoRI/NotI-cDNA adapter"
 - /number= 1

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION:1656..1669
- (D) OTHER INFORMATION:/product= "EcoRI/NotI-cDNA adapter"
 - /number= 2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GAATTCGCGG CCGCCTGAGA AACCTCGAA TTCAATTTTCG CATTGGCAG AG ATG
Met
1

55

AGA GGG AAC AAG TTT TGC TTT GAT TTA CGG TAC CTT CTC GTC GTG GCT
Arg Gly Asn Lys Phe Cys Phe Asp Leu Arg Tyr Leu Leu Val Val Ala
5 10 15

103

GCT CTC GCC TTC ATC TAC ATA CAG ATG CGG CTT TTC GCG ACA CAG TCA
Ala Leu Ala Phe Ile Tyr Ile Gln Met Arg Leu Phe Ala Thr Gln Ser
151

151

20 25 30 199
GAA TAT GTA GAC CGC CTT GCT GCT GCA ATT GAA GCA GAA AAT CAT TGT
Glu Tyr Val Asp Arg Leu Ala Ala Ala Ile Glu Ala Glu Asn His Cys
35 40 45
247
ACA AGT CAG ACC AGA TTG CTT ATT GAC AAG ATT AGC CAG CAG CAA GGA
Thr Ser Gln Thr Arg Leu Leu Ile Asp Lys Ile Ser Gln Gln Gln Gly
50 55 60 65
295
AGA GTA GTA GCT CTT GAA GAA CAA ATG AAG CAT CAG GAC CAG GAG TGC
Arg Val Val Ala Leu Glu Glu Gln Met Lys His Gln Asp Gln Glu Cys
70 75 80
343
CGG CAA TTA AGG GCT CTT GTT CAG GAT CTT GAA AGT AAG GGC ATA AAA
Arg Gln Leu Arg Ala Leu Val Gln Asp Leu Glu Ser Lys Gly Ile Lys
85 90 95
391
AAG TTA ATC GGA GAT GTG CAG ATG CCA GTG GCA GCT GTA GTT GTT ATG
Lys Leu Ile Gly Asp Val Gln Met Pro Val Ala Ala Val Val Val Met
100 105 110
439
GCT TGC AGT CGT ACT GAC TAC CTG GAG AGG ACT ATT AAA TCC ATC TTA
Ala Cys Ser Arg Thr Asp Tyr Leu Glu Arg Thr Ile Lys Ser Ile Leu
115 120 125
487
AAA TAC CAA ACA TCT GTT GCA TCA AAA TAT CCT CTT TTC ATA TCC CAG
Lys Tyr Gln Thr Ser Val Ala Ser Lys Tyr Pro Leu Phe Ile Ser Gln
130 135 140 145
535
GAT GGA TCA AAT CCT GAT GTA AGA AAG CTT GCT TTG AGC TAT GGT CAG
Asp Gly Ser Asn Pro Asp Val Arg Lys Leu Ala Leu Ser Tyr Gly Gln
150 155 160
583
CTG ACG TAT ATG CAG CAC TTG GAT TAT GAA CCT GTG CAT ACT GAA AGA
Leu Thr Tyr Met Gln His Leu Asp Tyr Glu Pro Val His Thr Glu Arg
165 170 175
631
CCA GGG GAA CTG GTT GCA TAC TAC AAG ATT GCA CGT CAT TAC AAG TGG
Pro Gly Glu Leu Val Ala Tyr Tyr Lys Ile Ala Arg His Tyr Lys Trp
180 185 190
679
GCA TTG GAT CAG CTG TTT CAC AAG CAT AAT TTT AGC CGT GTT ATC ATA
Ala Leu Asp Gln Leu Phe His Lys His Asn Phe Ser Arg Val Ile Ile
195 200 205
727
CTA GAA GAT GAT ATG GAA ATT GCT GCT GAT TTT TTT GAC TAT TTT GAG
Leu Glu Asp Asp Met Glu Ile Ala Ala Asp Phe Phe Asp Tyr Phe Glu
210 215 220 225
775
GCT GGA GCT ACT CTT CTT GAC AGA GAC AAG TCG ATT ATG GCT ATT TCT
Ala Gly Ala Thr Leu Leu Asp Arg Asp Lys Ser Ile Met Ala Ile Ser
230 235 240
823
TCT TGG AAT GAC AAT GGA CAA AGG CAG TTC GTC CAA GAT CCT GAT GCT
Ser Trp Asn Asp Asn Gly Gln Arg Gln Phe Val Gln Asp Pro Asp Ala
245 250 255

CTT TAC CGC TCA GAC TTT TTT CCT GGT CTT GGA TGG ATG CTT TCA AAA Leu Tyr Arg Ser Asp Phe Phe Pro Gly Leu Gly Trp Met Leu Ser Lys 260 265 270	871
TCA ACT TGG TCC GAA CTA TCT CCA AAG TGG CCA AAG GCT TAC TGG GAT Ser Thr Trp Ser Glu Leu Ser Pro Lys Trp Pro Lys Ala Tyr Trp Asp 275 280 285	919
GAC TGG CTA AGG CTG AAA GAA AAT CAC AGA GGT CGA CAA TTT ATT CGC Asp Trp Leu Arg Leu Lys Glu Asn His Arg Gly Arg Gln Phe Ile Arg 290 295 300 305	967
CCA GAA GTT TGC AGA ACG TAC AAT TTT GGT GAG CAT GGT TCT AGT TTG Pro Glu Val Cys Arg Thr Tyr Asn Phe Gly Glu His Gly Ser Ser Leu 310 315 320	1015
GGG CAG TTT TTT AAG CAG TAT CTT GAG CCA ATT AAG CTA AAT GAT GTC Gly Gln Phe Phe Lys Gln Tyr Leu Glu Pro Ile Lys Leu Asn Asp Val 325 330 335	1063
CAG GTT GAT TGG AAG TCA ATG GAC CTA AGT TAC CTT TTG GAG GAC AAC Gln Val Asp Trp Lys Ser Met Asp Leu Ser Tyr Leu Leu Glu Asp Asn 340 345 350	1111
TAT GTG AAA CAC TTT GGC GAC TTG GTT AAA AAG GCT AAG CCC ATC CAC Tyr Val Lys His Phe Gly Asp Leu Val Lys Lys Ala Lys Pro Ile His 355 360 365	1159
GGA GCT GAT GCT GTT TTG AAA GCA TTT AAC ATA GAT GGT GAT GTG CGT Gly Ala Asp Ala Val Leu Lys Ala Phe Asn Ile Asp Gly Asp Val Arg 370 375 380 385	1207
ATT CAG TAC AGA GAC CAA CTA GAC TTT GAA GAT ATC GCT CGA CAG TTT Ile Gln Tyr Arg Asp Gln Leu Asp Phe Glu Asp Ile Ala Arg Gln Phe 390 395 400	1255
GGC ATT TTT GAA GAA TGG AAG GAT GGT GTA CCA CGG GCA GCA TAT AAA Gly Ile Phe Glu Glu Trp Lys Asp Gly Val Pro Arg Ala Ala Tyr Lys 405 410 415	1303
GGG ATA GTA GTT TTC CGG TTT CAA ACA TCT AGA CGT GTG TTC CTT GTT Gly Ile Val Val Phe Arg Phe Gln Thr Ser Arg Arg Val Phe Leu Val 420 425 430	1351
TCC CCT GAT TCT CTT CGA CAA CTT GGA GTT GAA GAT ACT TAG Ser Pro Asp Ser Leu Arg Gln Leu Gly Val Glu Asp Thr * 435 440 445	1393
CGAAGATATG ATTGGAGCCT GAGCAACAAT TTAGACTTAT TTGGTAGGAT ACATTTGAAA	1453
GAGCTGACAC GAAAAGTATG ACTACCAGTA GCTACATGCA ACATTTTAAAT GTTAATGGAA	1513
GGAACCCACT GCTTATTGTT GGAATGGATG AATCATCACC ACATCCTATT ATTCAAGTTT	1573
ACAAACATAA AGAGGAAATG TTGCCCTATA AAAACAAATT TTTTGTCTTCT AAGAAGGAAC	1633
GTTACGATTA TGAGCAACTT TGGCGGCCGC GAATTC	1669

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 447 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

(11) MOBILE
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1

Met Arg Gly Asn Lys Phe Cys Phe Asp Leu Arg Tyr Leu Leu Val Val 15
1 5

Ala Ala Leu Ala Phe Ile Tyr Ile Gln Met Arg Leu Phe Ala Thr Gln 30
20 25

Ser Glu Tyr Val Asp Arg Leu Ala Ala Ile Glu Ala Glu Asn His 45
35 40

Cys Thr Ser Gln Thr Arg Leu Leu Ile Asp Lys Ile Ser Gln Gln Gln 60
50 55

Gly Arg Val Val Ala Leu Glu Glu Gln Met Lys His Gln Asp Gln Glu 80
65 70 75

Cys Arg Gln Leu Arg Ala Leu Val Gln Asp Leu Glu Ser Lys Gly Ile 95
85 90

Lys Lys Leu Ile Gly Asp Val Gln Met Pro Val Ala Ala Val Val Val 110
100 105

Met Ala Cys Ser Arg Thr Asp Tyr Leu Glu Arg Thr Ile Lys Ser Ile 125
115 120

Leu Lys Tyr Gln Thr Ser Val Ala Ser Lys Tyr Pro Leu Phe Ile Ser 140
130 135

Gln Asp Gly Ser Asn Pro Asp Val Arg Lys Leu Ala Leu Ser Tyr Gly 160
145 150 155

Gln Leu Thr Tyr Met Gln His Leu Asp Tyr Glu Pro Val His Thr Glu 175
165 170

Arg Pro Gly Glu Leu Val Ala Tyr Tyr Lys Ile Ala Arg His Tyr Lys 190
180 185

Trp Ala Leu Asp Gln Leu Phe His Lys His Asn Phe Ser Arg Val Ile 205
195 200

Ile Leu Glu Asp Asp Met Glu Ile Ala Ala Asp Phe Phe Asp Tyr Phe 220
210 215

Glu Ala Gly Ala Thr Leu Leu Asp Arg Asp Lys Ser Ile Met Ala Ile 240
225 230 235

Ser Ser Trp Asn Asp Asn Gly Gln Arg Gln Phe Val Gln Asp Pro Asp 255
245 250

Ala Leu Tyr Arg Ser Asp Phe Phe Pro Gly Leu Gly Trp Met Leu Ser 270
260 265

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(1) LENGTH: 1737 base pa

- SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1737 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
ORGANISM: N

- ORIGINAL SOURCE:
(A) ORGANISM: *Nicotiana tabacum*
(B) STRAIN: Samsun NN
(D) DEVELOPMENTAL STAGE: Sink organ
(F) TISSUE TYPE: Mesophyll
(G) CELL TYPE: Leaf cells

(vii) IMMEDIATE SOURCE:
LIBRARY: Lam

- IMMEDIATE SOURCE:
(A) LIBRARY: Lambda ZAP II (Eco RI)
(B) CLONE: gntI-A9(T)

URE:
(A) NAME/KEY: misc feature
733 741

(D) OTHER INFORMATION: /function= "Asn codon
is a potential glycosylation site"
consensus sequenc

```
/product= "N-glycosylation consensus  
type= "N-glycans modulate protein
```

```

/phenotype="N-glycosylation site"
properties="N-glycosylation site"
/standard_name="N-glycosylation site"

```

```

/standard_name= "N-glycosylation"
/label= pot-CHO
/note= "GnTI sequences from animals do not contain this

```

```

/note=
feature"

```

FEATURE:
(A) NAME/KEY: CDS
SECTION: 127

(A) NAME/REF: 350
(B) LOCATION: 127..1467
(C) IDENTIFICATION METHOD: experimental
(D) INFORMATION: /codon_start= 127

(B) LOCATION: 127-1

(C) IDENTIFICATION METHOD: experimental

(D) INFORMATION: /codon_start= 127

```

/function= "initiates con
secretory glycoproteins"

```

secretory glycoprotein
/EC number= 2.4.1.101

```
/EC_number=
/product=
```

```

/EC_number=2.4.1.101
/product=
"beta-1,2-N-acetylglucosaminyltransferase I"
/evidence= EXPERIMENTAL

```

"beta-1,2-N-acetyl-19-
/evidence= EXPERIMENTAL

```
/evidence=
/gene= "cgl"
```

```
/gene= "cgl"  
/standard_name= "gntI"
```

```
/standard=ORF
```

```

/standard_name= "gntI"
/label= ORF
/note= "first gntI sequence from tobacco (unpublished)"

```

FEATURE:
(A) NAME/KEY: 5'UTR
REGION: 15 12

(A) NAME/KEY: 501R
(B) LOCATION: 15..126

FEATURE:
(A) NAME/KEY: 3'UTR

(A) NAME/KEY: 3'0TR
(B) LOCATION: 1468..1723

FEATURE:
(A) NAME/KEY: CDS

(A) NAME/KEY: CDS
(B) LOCATION: 154..213

(A) NAME/KEY: CDS
(B) LOCATION: 154..213
(D) OTHER INFORMATION: /function= "membrane anchor (amino acids 10-29)" hydrophobic amino acid stretch in GntI type II

HER INFORMATION
acids 10-29)"

```

OTHER INFORMATION: /unclassified
acids 10-29)"
/product= "hydrophobic amino acid stretch in GnTI"
/feature_name= "membrane anchor of a type II

```

```

/product= "hydrophobic amino acid stretch in
/standard_name= "membrane anchor of a type II
golgi protein"

```

golgi protein"

FEATURE:
(A) NAME/KEY: misc_feature

(A) NAME/KEY: KRE-
(B) LOCATION: 1..14
INFORMAT

(A) NAME/KEY: misc_feature
(B) LOCATION: 1..14
(D) OTHER INFORMATION: /function= "use for cloning the
CDNA library in Lambda ZAPII"
-/-NotI-cDNA adapter"

OTHER INFORMATION: /function
CDNA library in Lambda ZAPII"
-I/NotI-CDNA ad

cDNA library in Lambda ZAP II
/product= "EcoRI/NotI-cDNA adapter"

```
/product=
/number= 1
```

FEATURE:
(A) NAME/KEY: misc_feature

(A) NAME/KEY: misc_feat
(B) LOCATION: 1724..1737
INFORMATION: /

FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 1724..1737
- (D) OTHER INFORMATION: /product= "EcoRI/NotI-cDNA adapter"
/number= 2

/number= 2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:	60
GAATTCGCGG CCGCCATTGA CTTGATCCTA ACTGAACAGG CAAAGTAAAT CCAGCGATGA	120
AGGCTAT TCGCTTCTC CTAAAGCCTT CAATCGAATT	180

GAATTCGCGG CCGCCATTGA CTTGATCCTA ACTGAACAGG
AACACTCATA ACTGAACACT GAGAGACTAT TCGCTTCTC CTAAAGCCTT CAATCGAATT

TTC CGG TAC CTC CTC
Leu Leu

AAACATCATA ACTGAACACT GAGAGACTAT TCGCTTCT
CGCAGC ATG AGA GGG AAC AAG TTT TGC TGT GAT TTC CGG TAC CTC CTC
Met Arg Gly Asn Lys Phe Cys Cys Asp Phe Arg Tyr Leu Leu
450 455 460
CTC GCG CTT TTT GCG

Met Arg Gyr 450 455 216
ATC TTG GCT GCT GTC GCC TTC ATC TAC ACA CAG ATG CGG CTT TTT GCG
Ile Leu Ala Ala Val Ala Phe Ile Tyr Thr Gln Met Arg Leu Phe Ala 475
465 470 264
ATT GAA GCA GAA

Ile Leu Ala Ala Val
465
ACA CAG TCA GAA TAT GCA GAT CGC CTT GCT GCT GCA ATT GAA GCA GAA
Thr Gln Ser Glu Tyr Ala Asp Arg Leu Ala Ala Ala Ile Glu Ala Glu
480
485
490
312

Thr Gln Ser Glu Tyr Lys Val Ile Met Phe Leu Thr
480

AAT CAT TGT ACA AGC CAG ACC AGA TTG CTT ATT GAC CAG ATT AGC CTG
Asn His Cys Thr Ser Gln Thr Arg Leu Leu Ile Asp Gln Ile Ser Leu
495

GAA GGA ATG AAG CGT CAG GAC
360

Asn His Cys Thr Ser 500 360
495
CAG CAA GGA AGA ATA GTT GCT CTT GAA GAA CAA ATG AAG CGT CAG GAC
Gln Gln Gly Arg Ile Val Ala Leu Glu Glu Gln Met Lys Arg Gln Asp
510 515 520 525
400
CCT GCT GCT CAG GAT CTT GAA AGT AAG
Leu Ser Lys

Gln Gln Gly Arg 115 515
510
CAG GAG TGC CGA CAA TTA AGG GCT CTT GTT CAG GAT CTT GAA AGT AAG
Gln Glu Cys Arg Gln Leu Arg Ala Leu Val Gln Asp Leu Glu Ser Lys 540
530 535
GTC GCT GCT GTA 45

Gln Glu Cys Arg Gln 530 535 456
GGC ATA AAA AAG TTG ATC GGA AAT GTA CAG ATG CCA GTG GCT GCT GTA
Gly Ile Lys Lys Leu Ile Gly Asn Val Gln Met Pro Val Ala Ala Val 555
545 550 504

Gly Ile Lys Lys 545 550 504
GTT GTT ATG GCT TGC AAT CGG GCT GAT TAC CTG GAA AAG ACT ATT AAA
Val Val Met Ala Cys Asn Arg Ala Asp Tyr Leu Glu Lys Thr Ile Lys 570 552

[illegible]

Ser Ile Leu Lys Val 580
 575
 ATA TCC CAG GAT GGA TCA CAT CCT GAT GTC AGG AAG CTT GCT TTG AGC
 Ile Ser Gln Asp Gly Ser His Pro Asp Val Arg Lys Leu Ala Leu Ser 600
 590
 595
 648

11e Ser Gln Asp Cys 595
590
TAT GAT CAG CTG ACG TAT ATG CAG CAC TTG GAT TTT GAA CCT GTG CAT
Tyr Asp Gln Leu Thr Met Gln His Leu Asp Phe Glu Pro Val His 620
610
615
690

Tyr Asp Gln Leu 610
 ACT GAA AGA CCA GGG GAG CTG ATT GCA TAC TAC AAA ATT GCA CGT CAT 696
 Thr Glu Arg Pro Gly Glu Leu Ile Ala Tyr Tyr Lys Ile Ala Arg His 635
 625
 630
 744

Thr Glu Arg Pro Gly Ser 630
625
TAC AAG TGG GCA TTG GAT CAG CTG TTT TAC AAG CAT AAT TTT AGC CGT 744
Tyr Lys Trp Ala Leu Asp Gln Leu Phe Tyr Lys His Asn Phe Ser Arg 650
640
645
650
655
660
665
670
675
680
685
690
695
700
705
710
715
720
725
730
735
740
745
750
755
760
765
770
775
780
785
790
795
800
805
810
815
820
825
830
835
840
845
850
855
860
865
870
875
880
885
890
895
900
905
910
915
920
925
930
935
940
945
950
955
960
965
970
975
980
985
990
995
1000
1005
1010
1015
1020
1025
1030
1035
1040
1045
1050
1055
1060
1065
1070
1075
1080
1085
1090
1095
1100
1105
1110
1115
1120
1125
1130
1135
1140
1145
1150
1155
1160
1165
1170
1175
1180
1185
1190
1195
1200
1205
1210
1215
1220
1225
1230
1235
1240
1245
1250
1255
1260
1265
1270
1275
1280
1285
1290
1295
1300
1305
1310
1315
1320
1325
1330
1335
1340
1345
1350
1355
1360
1365
1370
1375
1380
1385
1390
1395
1400
1405
1410
1415
1420
1425
1430
1435
1440
1445
1450
1455
1460
1465
1470
1475
1480
1485
1490
1495
1500
1505
1510
1515
1520
1525
1530
1535
1540
1545
1550
1555
1560
1565
1570
1575
1580
1585
1590
1595
1600
1605
1610
1615
1620
1625
1630
1635
1640
1645
1650
1655
1660
1665
1670
1675
1680
1685
1690
1695
1700
1705
1710
1715
1720
1725
1730
1735
1740
1745
1750
1755
1760
1765
1770
1775
1780
1785
1790
1795
1800
1805
1810
1815
1820
1825
1830
1835
1840
1845
1850
1855
1860
1865
1870
1875
1880
1885
1890
1895
1900
1905
1910
1915
1920
1925
1930
1935
1940
1945
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
2025
2030
2035
2040
2045
2050
2055
2060
2065
2070
2075
2080
2085
2090
2095
2100
2105
2110
2115
2120
2125
2130
2135
2140
2145
2150
2155
2160
2165
2170
2175
2180
2185
2190
2195
2200
2205
2210
2215
2220
2225
2230
2235
2240
2245
2250
2255
2260
2265
2270
2275
2280
2285
2290
2295
2300
2305
2310
2315
2320
2325
2330
2335
2340
2345
2350
2355
2360
2365
2370
2375
2380
2385
2390
2395
2400
2405
2410
2415
2420
2425
2430
2435
2440
2445
2450
2455
2460
2465
2470
2475
2480
2485
2490
2495
2500
2505
2510
2515
2520
2525
2530
2535
2540
2545
2550
2555
2560
2565
2570
2575
2580
2585
2590
2595
2600
2605
2610
2615
2620
2625
2630
2635
2640
2645
2650
2655
2660
2665
2670
2675
2680
2685
2690
2695
2700
2705
2710
2715
2720
2725
2730
2735
2740
2745
2750
2755
2760
2765
2770
2775
2780
2785
2790
2795
2800
2805
2810
2815
2820
2825
2830
2835
2840
2845
2850
2855
2860
2865
2870
2875
2880
2885
2890
2895
2900
2905
2910
2915
2920
2925
2930
2935
2940
2945
2950
2955
2960
2965
2970
2975
2980
2985
2990
2995
3000
3005
3010
3015
3020
3025
3030
3035
3040
3045
3050
3055
3060
3065
3070
3075
3080
3085
3090
3095
3100
3105
3110
3115
3120
3125
3130
3135
3140
3145
3150
3155
3160
3165
3170
3175
3180
3185
3190
3195
3200
3205
3210
3215
3220
3225
3230
3235
3240
3245
3250
3255
3260
3265
3270
3275
3280
3285
3290
3295
3300
3305
3310
3315
3320
3325
3330
3335
3340
3345
3350
3355
3360
3365
3370
3375
3380
3385
3390
3395
3400
3405
3410
3415
3420
3425
3430
3435
3440
3445
3450
3455
3460
3465
3470
3475
3480
3485
3490
3495
3500
3505
3510
3515
3520
3525
3530
3535
3540
3545
3550
3555
3560
3565
3570
3575
3580
3585
3590
3595
3600
3605
3610
3615
3620
3625
3630
3635
3640
3645
3650
3655
3660
3665
3670
3675
3680
3685
3690
3695
3700
3705
3710
3715
3720
3725
3730
3735
3740
3745
3750
3755
3760
3765
3770
3775
3780
3785
3790
3795
3800
3805
3810
3815
3820
3825
3830
3835
3840
3845
3850
3855
3860
3865
3870
3875
3880
3885
3890
3895
3900
3905
3910
3915
3920
3925
3930
3935
3940
3945
3950
3955
3960
3965
3970
3975
3980
3985
3990
3995
4000
4005
4010
4015
4020
4025
4030
4035

Tyr Lys Trp Ala Leu Met 645
640
GTT ATC ATA CTA GAA GAT GAT ATG GAA ATT GCC CCT GAT TTT TTT GAC 792
Val Ile Ile Leu Glu Asp Asp Met Glu Ile Ala Pro Asp Phe Phe Asp
655
660
GAG GAG AGA GAC AAG TCG ATT ATG 840

Val Ile Ile Leu Glu Asp 660
655
TTT TTT GAG GCT GGA GCT ACT CTT CTT GAC AGA GAC AAG TCG ATT ATG 840

Phe Phe Glu Ala Gly Ala Thr Leu Leu Asp Arg Asp Lys Ser Ile Met 685
 670 675
 GCT ATT TCT TCT TGG AAT GAC AAT GGA CAA ATG CAG TTT GTC CAA GAT 888
 Ala Ile Ser Ser Trp Asn Asp Asn Gly Gln Met Gln Phe Val Gln Asp 700
 690
 CCT TAT GCT CTT TAC CGC TCA GAT TTT TTT CCC GGT CTT GGA TGG ATG 936
 Pro Tyr Ala Leu Tyr Arg Ser Asp Phe Phe Pro Gly Leu Gly Trp Met 715
 705
 CTT TCA AAA TCT ACT TGG GAC GAA TTA TCT CCA AAG TGG CCA AAG GCT 984
 Leu Ser Lys Ser Thr Trp Asp Glu Leu Ser Pro Lys Trp Pro Lys Ala 730
 720
 TAC TGG GAC GAC TGG CTA AGA CTC AAA GAG AAT CAC AGA GGT CGA CAA 1032
 Tyr Trp Asp Asp Trp Leu Arg Leu Lys Glu Asn His Arg Gly Arg Gln 745
 735
 TTT ATT CGC CCA GAA GTT TGC AGA ACA TAT AAT TTT GGT GAG CAT GGT 1080
 Phe Ile Arg Pro Glu Val Cys Arg Thr Tyr Asn Phe Gly Glu His Gly 765
 750
 TCT AGT TTG GGG CAG TTT TTC AAG CAG TAT CTT GAG CCA ATT AAA CTA 1128
 Ser Ser Leu Gly Gln Phe Phe Lys Gln Tyr Leu Glu Pro Ile Lys Leu 780
 770
 AAT GAT GTC CAG GTT GAT TGG AAG TCA ATG GAC CTT AGT TAC CTT TTG 1176
 Asn Asp Val Gln Val Asp Trp Lys Ser Met Asp Leu Ser Tyr Leu Leu 795
 785
 GAG GAC AAT TAC GTG AAA CAC TTT GGT GAC TTG GTT AAA AAG GCT AAG 1224
 Glu Asp Asn Tyr Val Lys His Phe Gly Asp Leu Val Lys Lys Ala Lys 810
 800
 CCC ATC CAT GGA GCT GAT GCT GTC TTG AAA GCA TTT AAC ATA GAT GGT 1272
 Pro Ile His Gly Ala Asp Ala Val Leu Lys Ala Phe Asn Ile Asp Gly 825
 815
 GAT GTG CGT ATT CAG TAC AGA GAT CAA CTA GAC TTT GAA AAT ATC GCA 1320
 Asp Val Arg Ile Gln Tyr Arg Asp Gln Leu Asp Phe Glu Asn Ile Ala 845
 830
 CGG CAA TTT GGC ATT TTT GAA GAA TGG AAG GAT GGT GTA CCA CGT GCA 1368
 Arg Gln Phe Gly Ile Phe Glu Glu Trp Lys Asp Gly Val Pro Arg Ala 860
 850
 GCA TAT AAA GGA ATA GTA GTT TTC CGG TAC CAA ACG TCC AGA CGT GTA 1416
 Ala Tyr Lys Gly Ile Val Val Phe Arg Tyr Gln Thr Ser Arg Arg Val 875
 865
 TTC CTT GTT GGC CAT GAT TCG CTT CAA CAA CTC GGA ATT GAA GAT ACT 1464
 Phe Leu Val Gly His Asp Ser Leu Gln Gln Leu Gly Ile Glu Asp Thr 890
 880
 TAA CAAAGATATG ATTGCAGGAG CCCGGGCAAA ATTTTGTACT TATTGGGTAG 1517
 *
 GATGCATCGA GCTGACACTA AACCATGATT TTACCAGTTA CATAACAGT TTTAATGTTA 1577
 TACGGAGGAG CTCACGTGTT TAGTGTTGAA GGGATATCGG CTTCTTAGTA TTGGATGAAT 1637

1737

(2) INFORMATION:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 447 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

(11) MOLLE
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:

Met Arg Gly Asn Lys Phe Cys Cys Asp Phe Arg Tyr Leu Leu Ile Leu 15
1 5 10

Ala Ala Val Ala Phe Ile Tyr Thr Gln Met Arg Leu Phe Ala Thr Gln 30
20 25

Ser Glu Tyr Ala Asp Arg Leu Ala Ala Ala Ile Glu Ala Glu Asn His 45
35 40

Cys Thr Ser Gln Thr Arg Leu Leu Ile Asp Gln Ile Ser Leu Gln Gln 60
50 55

Gly Arg Ile Val Ala Leu Glu Glu Gln Met Lys Arg Gln Asp Gln Glu 80
65 70 75

Cys Arg Gln Leu Arg Ala Leu Val Gln Asp Leu Glu Ser Lys Gly Ile 95
85 90

Lys Lys Leu Ile Gly Asn Val Gln Met Pro Val Ala Ala Val Val Val 110
100 105

Met Ala Cys Asn Arg Ala Asp Tyr Leu Glu Lys Thr Ile Lys Ser Ile 125
115 120

Leu Lys Tyr Gln Ile Ser Val Ala Ser Lys Tyr Pro Leu Phe Ile Ser 140
130 135

Gln Asp Gly Ser His Pro Asp Val Arg Lys Leu Ala Leu Ser Tyr Asp 160
145 150 155

Gln Leu Thr Tyr Met Gln His Leu Asp Phe Glu Pro Val His Thr Glu 175
165 170

Arg Pro Gly Glu Leu Ile Ala Tyr Tyr Lys Ile Ala Arg His Tyr Lys 190
180 185

Trp Ala Leu Asp Gln Leu Phe Tyr Lys His Asn Phe Ser Arg Val Ile 205
195 200

Ile Leu Glu Asp Asp Met Glu Ile Ala Pro Asp Phe Phe Asp Phe Phe 220
210 215

Glu Ala Gly Ala Thr Leu Leu Asp Arg Asp Lys Ser Ile Met Ala Ile 240
225 230 235

Ser Ser Trp Asn Asp Asn Gly Gln Met Gln Phe Val Gln Asp Pro Tyr 255
245 250

[illegible]

Ala Leu Tyr Arg Ser Asp Phe Phe Pro Gly Leu Gly Trp Met Leu Ser
260 265 270

Lys Ser Thr Trp Asp Glu Leu Ser Pro Lys Trp Pro Lys Ala Tyr Trp
275 280 285

Asp Asp Trp Leu Arg Leu Lys Glu Asn His Arg Gly Arg Gln Phe Ile
290 295 300

Arg Pro Glu Val Cys Arg Thr Tyr Asn Phe Gly Glu His Gly Ser Ser
305 310 315 320

Leu Gly Gln Phe Phe Lys Gln Tyr Leu Glu Pro Ile Lys Leu Asn Asp
325 330 335

Val Gln Val Asp Trp Lys Ser Met Asp Leu Ser Tyr Leu Leu Glu Asp
340 345 350

Asn Tyr Val Lys His Phe Gly Asp Leu Val Lys Lys Ala Lys Pro Ile
355 360 365

His Gly Ala Asp Ala Val Leu Lys Ala Phe Asn Ile Asp Gly Asp Val
370 375 380

Arg Ile Gln Tyr Arg Asp Gln Leu Asp Phe Glu Asn Ile Ala Arg Gln
385 390 395 400

Phe Gly Ile Phe Glu Glu Trp Lys Asp Gly Val Pro Arg Ala Ala Tyr
405 410 415

Lys Gly Ile Val Val Phe Arg Tyr Gln Thr Ser Arg Arg Val Phe Leu
420 425 430

Val Gly His Asp Ser Leu Gln Gln Leu Gly Ile Glu Asp Thr *
435 440 445

(2) INFORMATION FOR SEQ ID NO: 5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1854 base pairs
 - (B) TYPE: Nucleotide
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA to mRNA
- (iii) HYPOTHETICAL: No
- (iv) ANTI-SENSE: No
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Arabidopsis thaliana
 - (B) STRAIN: Columbia
 - (D) DEVELOPMENTAL STAGE: Mature plants
 - (F) TISSUE TYPE: All tissues
- (vii) IMMEDIATE SOURCE:
 - (A) LIBRARY: Lambda Uni-ZAP (EcoRI/XhoI) and
Lambda ACT (XhoI)
 - (B) CLONE: pBSK-Ara-GntI-full #8
- (ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION:1185..1193
- (D) OTHER INFORMATION:/function= "Asn Codon is a potential glycosylation site"
- /product= "Consensus sequence for N-glycosylation"
- /phenotype= "N glycans modulate protein characteristics"
- /standard_name= "N glycosylation site"
- /label= pot-CHO
- /note= "absent in animal GnTI sequences"

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION:135..1469
- (C) IDENTIFICATION METHOD: experimental
- (D) OTHER INFORMATION:/codon_start= 135
- /function= "initiates complex N glycans on secretory glycoproteins"
- /EC_number= 2.4.1.101
- /product= "beta-1,2-N-acetyl glucosaminyl transferase I"
- /evidence= EXPERIMENTAL
- /gene= "cgl"
- /standard_name= "gntI"
- /label= ORF
- /note= "first gntI sequence from Arabidopsis (unpublished)"

(ix) FEATURE:

- (A) NAME/KEY: 5'UTR
- (B) LOCATION:19..134

(ix) FEATURE:

- (A) NAME/KEY: 3'UTR
- (B) LOCATION:1470..1848

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION:157..215
- (D) OTHER INFORMATION:/function= "membrane anchor (amino acids 8-27)"
- /product= "hydrophobic amino-acid region in GnTI"
- /standard_name= "membrane anchor of a Type II Golgi protein"
- /note= "identified by comparison with animal GnTI sequences "

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION:1..18
- (D) OTHER INFORMATION:/function= "for preparation of a cDNA library in Lambda ACT"
- /product= "XhoI-cDNA-Adaptor"
- /number= 1

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION:1849..1854
- (D) OTHER INFORMATION:/product= "XhoI-cDNA-Adaptor"
- /number= 2

000000-000000

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CTCGAGGCCA	CGAAGGCCAC	CGTTTTTGTT	ATAACGAACG	ACACCGTTTC	AAACAAC TTC	60
CTTATTAGCT	AGCTCCCTCC	CGGCGGCAAA	CACCAGAAGA	TCCACCGCTT	TTGATCTGGT	120
		GGC AGG ATC TCG TGT	GAC TTG AGA TTT CTT CTC			170

Met Ala Arg 5
1
ATC CCG GCA GCT TTC ATG TTC ATC TAC ATC CAG ATG AGG CTT TTC CAG 218
Ile Pro Ala Ala Phe Met Phe Ile Tyr Ile Gln Met Arg Leu Phe Gln 25
15
20
26

AAC CAT TGC ACT AGT CAA ATG CGA GGC CTC ATA GAT GAA GTT AGC ATC
 Asn His Cys Thr Ser Gln Met Arg Gly Leu Ile Asp Glu Val Ser Ile 60
 362

Lys Gln Ser Arg Ile Val
 65
 GAA GAA CTT GTG CAG CTT AAG GAT CTA ATC CAG ACG TTT GAA AAA AAA
 Glu Glu Leu Val Gln Leu Lys Asp Leu Ile Gln Thr Phe Glu Lys Lys
 80
 85
 90
 95
 100
 105
 110
 115
 120
 125
 130
 135
 140
 145
 150
 155
 160
 165
 170
 175
 180
 185
 190
 195
 200
 205
 210
 215
 220
 225
 230
 235
 240
 245
 250
 255
 260
 265
 270
 275
 280
 285
 290
 295
 300
 305
 310
 315
 320
 325
 330
 335
 340
 345
 350
 355
 360
 365
 370
 375
 380
 385
 390
 395
 400
 405
 410
 415
 420
 425
 430
 435
 440
 445
 450
 455
 460
 465
 470
 475
 480
 485
 490
 495
 500
 505
 510
 515
 520
 525
 530
 535
 540
 545
 550
 555
 560
 565
 570
 575
 580
 585
 590
 595
 600
 605
 610
 615
 620
 625
 630
 635
 640
 645
 650
 655
 660
 665
 670
 675
 680
 685
 690
 695
 700
 705
 710
 715
 720
 725
 730
 735
 740
 745
 750
 755
 760
 765
 770
 775
 780
 785
 790
 795
 800
 805
 810
 815
 820
 825
 830
 835
 840
 845
 850
 855
 860
 865
 870
 875
 880
 885
 890
 895
 900
 905
 910
 915
 920
 925
 930
 935
 940
 945
 950
 955
 960
 965
 970
 975
 980
 985
 990
 995
 1000

Gly Ile Ala Lys Leu Thr Ser 100
 95
 GTG GTT ATG GCC TGC AGT CGT GCA GAC TAT CTT GAA AGG ACT GTT AAA
 Val Val Met Ala Cys Ser Arg Ala Asp Tyr Leu Glu Arg Thr Val Lys 120
 110
 GGT GCT TCA AAA TAT CCT CTA TTT
 115
 506
 554

TCA Val Leu Thr Tyr Gln 130
 Ser Val Leu Thr Tyr Gln 130
 125
 ATA TCT CAG GAT GGA TCT GAT CAA GCT GTC AAG AGC AAG TCA TTG AGC
 Ile Ser Gln Asp Gly Ser Asp Gln Ala Val Lys Ser Lys Ser Leu Ser
 145 150 155
 650

ACT GAA AGG CCT GGT GAA CTG ACT GCG TAC TAC AAG ATT GCA CGT CAC
Thr Glu Arg Pro Gly Glu Leu Thr Ala Tyr Tyr Lys Ile Ala Arg His

Tyr Lys Trp Ala Leu Asp 195
190
GTG ATT ATA CTA GAA GAC GAT ATG GAA ATT GCT CCA GAC TTC TTT GAT
Val Ile Ile Leu Glu Asp Asp Met Glu Ile Ala Pro Asp Phe Phe Asp 220
205 210 215

TAC TTT GAG GCT GCA GCT AGT CTC ATG GAT AGG GAT AAA ACC ATT ATG 842
Tyr Phe Glu Ala Ala Ala Ser Leu Met Asp Arg Asp Lys Thr Ile Met 235

GCT GCT TCA TCA TGG AAT GAT AAT GGA CAG AAG CAG TTT GTG CAT GAT 890
Ala Ala Ser Ser Trp Asn Asp Asn Gly Gln Lys Gln Phe Val His Asp 250

CCC TAT GCG CTA TAC CGA TCA GAT TTT TTT CCT GGC CTT GGG TGG ATG 938
Pro Tyr Ala Leu Tyr Arg Ser Asp Phe Phe Pro Gly Leu Gly Trp Met 265

CTC AAG AGA TCG ACT TGG GAT GAG TTA TCA CCA AAG TGG CCA AAG GCT 986
Leu Lys Arg Ser Thr Trp Asp Glu Leu Ser Pro Lys Trp Pro Lys Ala 280

TAC TGG GAT GAT TGG CTG AGA CTA AAG GAA AAC CAT AAA GGC CGC CAA 1034
Tyr Trp Asp Asp Trp Leu Arg Leu Lys Glu Asn His Lys Gly Arg Gln 300

TTC ATT GCA CCG GAA GTC TGT AGA ACA TAC AAT TTT GGT GAA CAT GGG 1082
Phe Ile Ala Pro Glu Val Cys Arg Thr Tyr Asn Phe Gly Glu His Gly 315

TCT AGT TTG GGA CAG TTT TTC AGT CAG TAT CTG GAA CCT ATA AAG CTA 1130
Ser Ser Leu Gly Gln Phe Phe Ser Gln Tyr Leu Glu Pro Ile Lys Leu 330

AAC GAT GTG ACG GTT GAC TGG AAA GCA AAG GAC CTG GGA TAC CTG ACA 1178
Asn Asp Val Thr Val Asp Trp Lys Ala Lys Asp Leu Gly Tyr Leu Thr 345

GAG GGA AAC TAT ACC AAG TAC TTT TCT GGC TTA GTG AGA CAA GCA CGA 1226
Glu Gly Asn Tyr Thr Lys Tyr Phe Ser Gly Leu Val Arg Gln Ala Arg 360

CCA ATT CAA GGT TCT GAC CTT GTC TTA AAG GCT CAA AAC ATA AAG GAT 1274
Pro Ile Gln Gly Ser Asp Leu Val Leu Lys Ala Gln Asn Ile Lys Asp 380

GAT GAT CGT ATC CGG TAT AAA GAC CAA GTA GAG TTT GAA CGC ATT GCA 1322
Asp Asp Arg Ile Arg Tyr Lys Asp Gln Val Glu Phe Glu Arg Ile Ala 395

GGG GAA TTT GGT ATA TTT GAA GAA TGG AAG GAT GGT GTG CCA CGA ACA 1370
Gly Glu Phe Gly Ile Phe Glu Glu Trp Lys Asp Gly Val Pro Arg Thr 410

GCA TAT AAA GGA GTA GTG GTG TTT CGA ATC CAG ACA ACA AGA CGT GTA 1418
Ala Tyr Lys Gly Val Val Phe Arg Ile Gln Thr Thr Arg Arg Val 425

TTC CTG GTT GGG CCA GAT TCT GTA ATG CAG CTT GGA ATT CGA AAT TCC 1466
Phe Leu Val Gly Pro Asp Ser Val Met Gln Leu Gly Ile Arg Asn Ser 440

TGA TGCAAAACAT ATGAAAGGAA AAGAAGATTT TGGACCGCAT GCAGCCTCCT 1519
* 445

TCTAGCAGCT GTTAGGTTGT ATTGTTATTT ATGGATGAGT TTGTAGAGCG GTGGGGTTAA 1579

CTTTAACAGC AAGGAAGCTC TGGTGACCAG GCTGATTGGC TTAGAAGTTA TGGAACCCC 1639

RECEIVED
JAN 10 1964
U.S. DEPARTMENT OF AGRICULTURE
WASHINGTON, D.C.

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 445 amino acids
(B) TYPE: Amino acid
(D) TOPOLOGY: Linear

(ii) MOLECULE TYPE: Protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

(ii) MOLECULE NAME:
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1

Met Ala Arg Ile Ser Cys Asp Leu Arg Phe Leu Leu Ile Pro Ala Ala
1 5 10 15
Phe Met Phe Ile Tyr Ile Gln Met Arg Leu Phe Gln Thr Gln Ser Gln
20 25 30
Tyr Ala Asp Arg Leu Ser Ser Ala Ile Glu Ser Glu Asn His Cys Thr
35 40 45
Ser Gln Met Arg Gly Leu Ile Asp Glu Val Ser Ile Lys Gln Ser Arg
50 55 60
Ile Val Ala Leu Glu Asp Met Lys Asn Arg Gln Asp Glu Glu Leu Val
65 70 75 80
Gln Leu Lys Asp Leu Ile Gln Thr Phe Glu Lys Lys Gly Ile Ala Lys
85 90 95
Leu Thr Gln Gly Gly Gln Met Pro Val Ala Ala Val Val Met Ala
100 105 110
Cys Ser Arg Ala Asp Tyr Leu Glu Arg Thr Val Lys Ser Val Leu Thr
115 120 125
Tyr Gln Thr Pro Val Ala Ser Lys Tyr Pro Leu Phe Ile Ser Gln Asp
130 135 140
Gly Ser Asp Gln Ala Val Lys Ser Lys Ser Leu Ser Tyr Asn Gln Leu
145 150 155 160
Thr Tyr Met Gln His Leu Asp Phe Glu Pro Val Val Thr Glu Arg Pro
165 170 175
Gly Glu Leu Thr Ala Tyr Tyr Lys Ile Ala Arg His Tyr Lys Trp Ala
180 185 190
Leu Asp Gln Leu Phe Tyr Lys His Lys Phe Ser Arg Val Ile Ile Leu
195 200 205
Glu Asp Asp Met Glu Ile Ala Pro Asp Phe Phe Asp Tyr Phe Glu Ala
210 215 220
Ala Ala Ser Leu Met Asp Arg Asp Lys Thr Ile Met Ala Ala Ser Ser

225 230 235 240

Trp Asn Asp Asn Gly Gln Lys Gln Phe Val His Asp Pro Tyr Ala Leu
245 250 255

Tyr Arg Ser Asp Phe Phe Pro Gly Leu Gly Trp Met Leu Lys Arg Ser
260 265 270

Thr Trp Asp Glu Leu Ser Pro Lys Trp Pro Lys Ala Tyr Trp Asp Asp
275 280 285

Trp Leu Arg Leu Lys Glu Asn His Lys Gly Arg Gln Phe Ile Ala Pro
290 295 300

Glu Val Cys Arg Thr Tyr Asn Phe Gly Glu His Gly Ser Ser Leu Gly
305 310 315 320

Gln Phe Phe Ser Gln Tyr Leu Glu Pro Ile Lys Leu Asn Asp Val Thr
325 330 335

Val Asp Trp Lys Ala Lys Asp Leu Gly Tyr Leu Thr Glu Gly Asn Tyr
340 345 350

Thr Lys Tyr Phe Ser Gly Leu Val Arg Gln Ala Arg Pro Ile Gln Gly
355 360 365

Ser Asp Leu Val Leu Lys Ala Gln Asn Ile Lys Asp Asp Arg Ile
370 375 380

Arg Tyr Lys Asp Gln Val Glu Phe Glu Arg Ile Ala Gly Glu Phe Gly
385 390 395 400

Ile Phe Glu Glu Trp Lys Asp Gly Val Pro Arg Thr Ala Tyr Lys Gly
405 410 415

Val Val Val Phe Arg Ile Gln Thr Thr Arg Arg Val Phe Leu Val Gly
420 425 430

Pro Asp Ser Val Met Gln Leu Gly Ile Arg Asn Ser *
435 440 445